

## **Executive Summary**

This executive summary provides a condensed overview of the final project "*Student Achievement in Secondary Education of Two Portuguese Schools*." The analysis delves into the myriad factors influencing student success in Mathematics and Portuguese, utilizing a diverse dataset comprising student grades, demographic details, social factors, and school-related features. The project aims to uncover the underlying determinants of academic achievement in the specified subjects through two primary tasks. Through systematic exploration, data preprocessing, and model development, the project seeks to gain valuable insights into the factors shaping performance outcomes in secondary education. The modeling phase entails implementing regression and classification models to address the defined tasks, followed by evaluation using metrics such as confusion matrix, ROC, and AUC curves.

### **Key findings:**

- **Demographic Aspects:** Gender disparities were observed, with females outperforming males in Portuguese, while males excelled in Mathematics. Additionally, students with parents living apart demonstrated better academic performance compared to those with parents living together.
- **School-Related Aspects:** Course preference emerged as the primary factor influencing students' school choices, with Gabriel Pereira students generally outperforming Mousinho da Silveira students in both subjects.
- **Social Aspects:** Positive family dynamics, moderate free time, limited socializing, and minimal alcohol consumption correlated with better academic outcomes. Involvement in romantic relationships showed a potential impact on academic focus and productivity.
- **Correlation Analysis:** Participation in extracurricular activities, aspiration for higher education, and internet access at home positively correlated with academic performance. Conversely, negative correlations were observed between absences, gender, alcohol consumption, travel time, academic failures, and grades in Portuguese.

### **Key Metrics or Data:**

The project evaluated various regression and classification models to predict student achievement in Mathematics and Portuguese, yielding insightful metrics for performance assessment.

For Task 1, regression models were employed, with the Decision Tree model exhibiting the highest R-squared value of approximately 0.83. This model also achieved the lowest mean absolute error (MAE) of about 0.47 and a mean squared error (MSE) of around 1.18, signifying superior predictive accuracy and minimal prediction errors.

In Task 2, classification models were utilized, where both the Decision Tree and Random Forest models displayed the highest accuracy of 76%. Additionally, these models demonstrated balanced performance across precision, recall, and F1-score metrics, with precision (weighted) values of 0.79 and 0.78, respectively.

These key metrics underscore the efficacy of the Decision Tree and Random Forest models in predicting student achievement in secondary education, providing valuable insights for academic interventions and support strategies.

### **Conclusion:**

The Decision Tree and Random Forest models emerged as robust predictors of student achievement in secondary education, providing educators with valuable insights for early intervention and support strategies. These models demonstrated proficiency in explaining variance in student performance and categorizing performance levels, enabling tailored interventions based on individual needs. By leveraging these insights, educators can proactively address academic challenges and foster student success in Mathematics and Portuguese.